

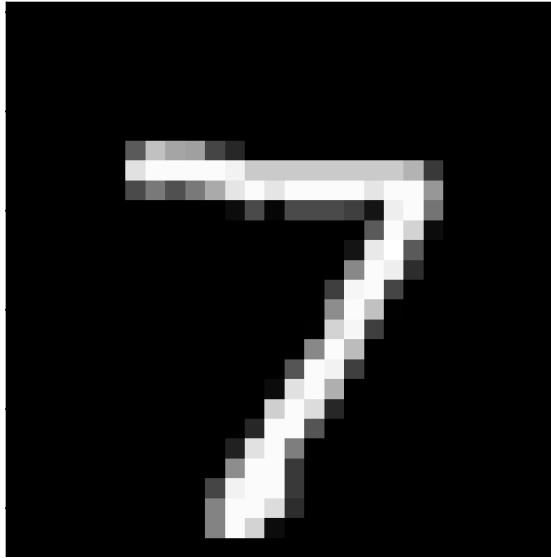
# Dataset

Ki Hyun Kim

[nlp.with.deep.learning@gmail.com](mailto:nlp.with.deep.learning@gmail.com)

# Our Objective

- 주어진 데이터에 대해서 결과를 내는 가상의 함수를 **모사**하는 함수를 만드는 것
  - e.g. 주어진 숫자 그림을 보고 숫자 맞추기



## In order to do this:

- 가상의 함수(ground-truth)를 통해 데이터 쌍( $x, y$ )을 수집해 데이터셋을 만들고, 우리의 함수(e.g. 신경망)가 데이터를 통해 가상의 함수를 모방하도록 한다.

$$D = \{(x_i, y_i)\}_{i=1}^N$$

- 이때,  $x$ 와  $y$ 는 각각  $n$ 차원과  $m$ 차원의 벡터로 표현될 수 있다.

$$x \in \mathbb{R}^n, y \in \mathbb{R}^m$$

# Example: Tabular Dataset

- 여러 column으로 이루어진 테이블로 구성된 데이터셋 (e.g. 엑셀 파일)
  - 한 row가 한 개의 sample을 의미함
  - Column 갯수가 벡터의 차원을 의미

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	TARGET
<b>501</b>	0.06263	0.0	11.93	0.0	0.573	6.593	69.1	2.4786	1.0	273.0	21.0	391.99	9.67	22.4
<b>502</b>	0.04527	0.0	11.93	0.0	0.573	6.120	76.7	2.2875	1.0	273.0	21.0	396.90	9.08	20.6
<b>503</b>	0.06076	0.0	11.93	0.0	0.573	6.976	91.0	2.1675	1.0	273.0	21.0	396.90	5.64	23.9
<b>504</b>	0.10959	0.0	11.93	0.0	0.573	6.794	89.3	2.3889	1.0	273.0	21.0	393.45	6.48	22.0
<b>505</b>	0.04741	0.0	11.93	0.0	0.573	6.030	80.8	2.5050	1.0	273.0	21.0	396.90	7.88	11.9